Cell

Supplemental Information

Structural Repertoire of HIV-1-Neutralizing Antibodies Targeting the CD4 Supersite in 14 Donors

Tongqing Zhou, Rebecca M. Lynch, Lei Chen, Priyamvada Acharya, Xueling Wu, Nicole A. Doria-Rose, M. Gordon Joyce, Daniel Lingwood, Cinque Soto, Robert T. Bailer, Michael J. Ernandes, Rui Kong, Nancy S. Longo, Mark K. Louder, Krisha McKee, Sijy O'Dell, Stephen D. Schmidt, Lillian Tran, Zhongjia Yang, Aliaksandr Druz, Timothy S. Luongo, Stephanie Moquin, Sanjay Srivatsan, Yongping Yang, Baoshan Zhang, Anqi Zheng, Marie Pancera, Tatsiana Kirys, Ivelin S. Georgiev, Tatyana Gindin, Hung-Pin Peng, An-Suei Yang, NISC Comparative Sequencing Program, James C. Mullikin, Matthew D. Gray, Leonidas Stamatatos, Dennis R. Burton, Wayne C. Koff, Myron S. Cohen, Barton F. Haynes, Joseph P. Casazza, Mark Connors, Davide Corti, Antonio Lanzavecchia, Quentin J. Sattentau, Robin A. Weiss, Anthony P. West, Jr. Pamela J. Bjorkman, Johannes F. Scheid, Michel C. Nussenzweig, Lawrence Shapiro, John R. Mascola, and Peter D. Kwong,

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Isolation and Expression of Antibody 44-VRC13.01, C38-VRC16.01, C38-VRC18.01, and Z258-VRC27.01

Antigen-specific memory B cells from donors NIH44, NIHC38 and NIHZ258 were isolated with Avi-tagged RSC3 and ΔRSC3 and single-cell sorted on a FACS Aria II as described previously (Wu et al., 2010). Kappa and lambda light chain gene and IgG heavy chain gene variable regions were amplified using nested PCR, and positive 2nd round PCR products were cherry-picked for direct sequencing. Heavy and light chain nucleotide sequences of the variable region were analyzed using IMGT\V-Quest (http://www.imgt.org/IMGT_vquest/share/textes/imgtvquest.html). First-round PCR products were re-amplified with custom primers containing restriction digest sites followed by subcloning, expression and purification as previously described (Wu et al., 2010). Heavy chains were reconstituted as IgG1. The antibody names are designated with a “donor-antibody lineage.clone” convention and therefore called 44-VRC13.01, C38-VRC16.01, C38-VRC18.01 and Z258-VRC27.01.

Neutralization Assays

Single round of replication Env-pseudoviruses were prepared, titered and used to infect TZM-bl target cells as described previously (Montefiori, 2009). Neutralization breadth of VRC13.01, VRC16.01, VRC18.01 and VRC27.01 were determined using a previously described panel of 176 geographically and genetically diverse Env-pseudoviruses representing the major subtypes and circulating recombinant forms (Wu et al., 2010). The data were calculated as a reduction in luminescence units compared with control wells, and reported as half-maximum inhibitory concentration (IC_{50}) in micrograms per milliliter for monoclonal antibodies.

X-Ray Crystallography

The gp120-antibody complexes were formed by mixing deglycosylated gp120 with the antibody Fab in a 1:2 molar ratio. The complexes were purified by size exclusion chromatography (Hiload 26/60 Superdex S200 prep grade; GE Healthcare) with buffer containing 0.35 M NaCl, 2.5 mM Tris (pH 7.5), and 0.02% NaN_3. Fractions with gp120-antibody complexes were concentrated to ~10 mg/ml and used for crystallization experiments. All gp120-Fab complex were screened against 576 crystallization conditions using a Cartesian Honeybee crystallization robot. Initial crystals were grown by the vapor diffusion method in sitting drops at 20 °C by mixing 0.2 µl of protein complex with 0.2 µl of reservoir solution. Crystals were manually reproduced in hanging drops by mixing 0.50 µl protein complex with 0.5 µl reservoir solution.

The HT593 gp120-HJ16 complex was crystallized with a reservoir solution containing 25% (w/v) PEG2000, 0.2 M Li_2SO_4, 0.1 M Tris-HCl pH 8.5 and 5% (v/v) isopropanol and was flash frozen in liquid nitrogen with 15% 2R, 3R-butanediol (v/v) as a cryoprotectant. The 93TH057 $core_e$ gp120-

VRC13 complex was crystallized with a reservoir solution of 9.5% w/v PEG 8000, 15% (v/v) isopropanol, 0.1 M imidazole, pH 6.5 and was flash frozen in liquid nitrogen with 15% (v/v) 2R,3R-butanediol as a cryoprotectant. The 93TH057 $core_e$ gp120-VRC16 complex was crystallized with a reservoir solution of 7.5% (w/v) PEG 8000, 0.1 M CHES, pH 9.5 and was flash frozen in liquid nitrogen with 30 % (v/v) glycerol as a cryoprotectant. The 93TH057 $core_e$ gp120-VRC18 complex was crystallized with a reservoir solution of 9.0% (w/v) PEG 8000, 4.5% (v/v) isopropanol, 0.1 M HEPES, pH 7.5 and was flash frozen in liquid nitrogen with 15% (v/v) 2R,3R-butanediol as a cryoprotectant. The Q23.17 $core_e$ gp120-VRC27 complex was crystallized with a reservoir solution of 4% (v/v) PEG 400, 1.9 M $(NH_4)_2SO_4$, 0.1 M Tris-HCl, pH 8.5 and was flash frozen in liquid nitrogen with 10% (v/v) 2R,3R-butanediol as a cryoprotectant. The 93TH057 $core_e$ gp120-1B2530 complex was crystallized with a reservoir solution of 8% w/v PEG 8000, 0.2 M sodium acetate and 0.1 m sodium cacodylate, pH 6.5 and was flash frozen in liquid nitrogen with 15%(v/v) 2R,3R-butanediol as a cryoprotectant. The YU2 $core_e$ gp120-8ANC131 complex was crystallized with a reservoir solution of 6% w/v PEG 3000, 0.1 M HEPES pH 7.5 and 0.2M NaCl, and was flash frozen in liquid nitrogen with 30% ethylene glycol as a cryoprotectant. The Q842.d12 $core_e$ gp120-8ANC134 complex was crystallized with a reservoir solution of 2.0% (v/v) PEG 400, 2.0 M $(NH_4)_2SO_4$ and was flash frozen in liquid nitrogen with 30% xylitol as a cryoprotectant.

Data for all crystals were collected at a wavelength of 1.00Å at SER-CAT beamlines ID-22 and BM-22 (Advanced Photon Source, Argonne National Laboratory). All diffraction data were processed with the HKL2000 suite (Otwinowski and Minor, 1997), structures were solved by molecular replacement using PHASER (McCoy et al., 2007), and iterative model building and refinement were performed in COOT (Emsley and Cowtan, 2004) and PHENIX (Adams et al., 2010), respectively. For HT593$core_e$-HJ16 complex, a molecular replacement solution was obtained with search model from PDB ID 3JWD. For 93TH057$core_e$ complexes with VRC13, VRC16 and VRC18, molecular replacement solutions were obtained using PDB ID 3SE9 as a search model. For Q23.17$core_e$-VRC27 complex, a molecular replacement solutions was obtained with search model from PDB ID 4J6R. For YU2$core_e$-8ANC131 complex, a molecular replacement solution consisting of one complex per asymmetric unit was obtained with search models from PDB ID 3TGQ (unliganded YU2 $core_e$) and PDB ID 3NGB (VRC01 Fab). For Q842d12$core_e$-8ANC134 complex, a molecular replacement solution consisting of two complexes per asymmetric unit was obtained using the YU2$core_e$-8ANC131 structure as search model. For 93TH057 $core_e$-1B2530 complex, data were initially processed in $P2_12_12_1$ space group. However, phenix.xtriage showed a multivariate Z score of 7.136. The data were reprocessed in $P2_1$ space group and molecular replacement solution consisting of two complexes per asymmetric unit was obtained using search models from PDB ID 4LSU, with a final Z-score of 22.2 and log-likelihood gain of 1219. Running Phenix.xtriage with the newly processed data suggested twinning. Phenix.twin_map_utils was used to determine twin law (twin_law=h,-k,-l) and twin fraction (0.50), which were used throughout refinement.

Cross validation (R_{free}) test sets consisting of 5% of the data were used throughout the refinement processes and hydrogen atoms were included in the refinement model. Structure validations were performed periodically during the model building/refinement process with MolProbity (Davis et al., 2004). The HT593 $core_e$-HJ16 structure was refined to a final R_{free} value of 21% with 93.2% residues in the favored region of the Ramachandran plot, and 1.1% outliers. The 93TH057 $core_e$-VRC13 structure was refined to a final R_{free} value of 24.9% with 92.0% residues in the favored region of the Ramachandran plot, and 1.4% outliers. The 93TH057 $core_e$-VRC16 structure was refined to a final R_{free} value of 20.4% with 95.8% residues in the favored region of the Ramachandran plot, and 0.39% outliers. The 93TH057 $core_e$-VRC18 structure was refined to a final R_{free} value of 20.9% with 98% residues in the favored region of the Ramachandran plot, and no outliers. The Q23.17$core_e$-VRC27 structure was refined to a final R_{free} value of 27.4% with 93.3% residues in the favored region of the Ramachandran plot, and 2.14% outliers. The 93TH057 $core_e$-1B2530 structure was refined to a final R_{free} value of 35% with 93.04% residues in the favored region of the Ramachandran plot, and 19 (1.25%) outliers.YU2$core_e$-8ANC131 structure was refined to a final R_{free} value of 25.2% with 96.3% residues in the favored region of the Ramachandran plot, and 3 (0.39%) outliers. The Q842d12$core_e$-8ANC134 structure was refined to a final R_{free} value of 26.4% with 96.2% residues in the favored region of the Ramachandran plot, and 7 (0.48%) outliers (Table S2). All figures containing representations of protein crystal structures were made with PyMOL (DeLano, 2002). gp120 and antibody interactions were analyzed with the PISA server (Krissinel and Henrick, 2007) (Table S6).

Neutralization Fingerprints

The neutralization fingerprint of an antibody is defined as the potency pattern with which the antibody neutralizes a set of viral strains. For a set of strains representing the global HIV-1 Env diversity, antibody neutralization fingerprints can be used as a predictor of the antibody epitopes: antibodies targeting similar epitopes exhibit similar fingerprints, and vice versa (Georgiev et al., 2013). Antibody neutralization data over a set of 168 diverse HIV-1 strains (the largest set of strains for which data was available for all antibodies included in the analysis) was used in the analysis. For antibodies with multiple repeats of the neutralization experiment, the average of the repeats was used in the computation. Neutralization-based antibody clustering was performed as described previously (Georgiev et al., 2013); briefly, Spearman correlations were computed between the neutralization fingerprints for each pair of antibodies, effectively transforming the antibody-virus neutralization matrix into an antibody-antibody correlation matrix. The computed antibody-antibody correlations were then used to cluster the antibodies according to neutralization fingerprint similarity. The computation was performed using Mathematica code as published previously (Georgiev et al., 2013).

Germline Gene Usage of the VH Gene-Restricted Antibodies

Two key features of the VH1-2- and VH1-46-restricted classes of CD4 binding site antibodies are the CDR H2-dominated interface and the Arg71 interaction with Asp368 of gp120. In these antibodies the conformation of CDR H2 is critical for interacting with the CD4 binding site loop of gp120 (Zhou et al., 2010). The conformation of CDR H2 in the VH1-2 and VH1-46 antibodies is in fact a very common conformation of this loop in antibody structures - designated the H2-10-1 conformation as defined by North and colleagues (North et al., 2011). Since this is a common conformation, it might not seem to be a critical factor in germline selection. However, the identity of residue 71 has been noted to strongly correlate with the conformation of CDR H2 (North et al., 2011; Tramontano et al., 1990). Specifically, when residue 71 is Arg, CDR H2 generally adopts a very different conformation (H2-10-2) and the side chain of Arg71 is buried between CDR H1 and CDR H2. Both the H2-10-2 conformation and the buried Arg71 would be incompatible with CD4-mimetic binding to gp120. Why does this not happen for the VH-restricted antibodies? One possibility is that Pro52A may stabilize the H2-10-1 (VRC01-like) conformation of CDR H2 and this correlates with the Arg71 side chain being exposed rather than buried. Of 29 of the most frequently utilized human VH genes, only VH1-2 and VH1-46 have both Arg71 and Pro52A. Only two families of human V genes possessing Arg71: VH1 and VH3. However, none of the functional germline VH3 family members (*01 alleles) have Pro52A.

We have examined structures of various Arg71-containing VH1- and VH3- derived antibodies to determine whether the side chain of Arg71 is in an exposed or buried position. Overall, VH3-derived antibodies generally bury Arg71 between CDR H1 and CDR H2. Conversely, the VH1-2- and VH1-46-derived antibody structures that we have examined (including 5 non-CD4 binding site antibodies) all have Arg71 exposed. Antibodies derived from VH1-3, which lacks Pro52A, also have Arg71 exposed. In this case residue 58 (a Lys in VH1-3) may play a role in preventing this gene from forming CD4 mimetic antibodies.

In addition, there are fine distinctions between CDR loop conformations (North et al., 2011). Strictly speaking, the CDR H2 loops of VH-restricted antibodies have H2-10-1-like conformations. The VRC01-gp120 complex structure (Zhou et al., 2010) has conformation H2-10-1, while the NIH45-46-gp120 complex structure (Diskin et al., 2011) has conformation H2-10-6, which differs by one peptide-flip (the amide bond plane connecting two adjacent C-alphas is flipped by 180 degrees). For our purposes there are two classes: H2-10-1-like (VRC01-like), or H2-10-2 (a very different conformation that also buries the Arg71 side chain). Some VH1-2 and VH1-46 class antibodies lack Pro52A - namely VRC-PG04 (with Thr52A) and 8ANC131/8ANC134 (residue 52A deleted). These changes presumably happened after initial selection, and it is possible that other residues changes may have occurred which stabilize the H2-10-1-like conformation independently of Pro52A.

It is worth noting that VH-restricted antibodies have been isolated from 11 individuals thus far. It is possible that compatible but infrequently-used germline genes may not have been observed.

Since antibody development is a stochastic process, we have restricted our analysis to the more frequently used VH genes – specifically, those 29 VH gene segments with a frequency >~1.1%, as measured by (Arnaout et al., 2011).Pro52A appears to favor the CDR H2 conformation necessary for CD4-mimetic binding and helps ensure that Arg71 is exposed rather than buried between CDR H1 and CDR H2 to promote the critical interaction with $Asp368_{gp120}$. Large residues at position 58 may be incompatible for VH-restricted antibodies due to steric clashes with V5 and Loop D.

High-Throughput Sequencing

Amplicon for 454 next-generation sequencing was prepared as described (Wu et al., 2011) (Zhu et al., 2012) with slight modifications as indicated. Briefly, mRNA was prepared from 10–20 million PBMC using an Oligotex kit (Qiagen). cDNA was synthesized using Superscript II reverse transcriptase (Invitrogen) and oligo-dT(12–18) primers. Individual PCR reactions were performed with Phusion polymerase for 30 cycles. Primers had adapters for 454 next generation sequencing and have all been previously published. Forward primers were either specific for all VH1 (Scheid et al., 2011) or VH3 (Tiller et al., 2008) family genes while reverse primers were a mix of mu and gamma specific (Doria-Rose et al., 2014). Pyrosequencing of the PCR products was performed on a GSFLX sequencing instrument (Roche-454 Life Sciences, Bradford, CT, USA) on a half chip per reaction. The averaged raw reads are listed in Table S6.

Bioinformatics Processing of 454 Data and Cross-Donor Phylogenetic Analysis

All six 454 heavy chain sets were filtered for read length keeping only those sequences with a minimum of 300 nucleotides. The remaining sequences from each data set were then submitted to the IMGT High-Vquest server (http://www.imgt.org/IMGTindex/IMGTHighV-QUEST.html) for processing and germline gene assignment. Each processed 454 data set was filtered using four VH family germline genes: 1) VH1-2; 2) VH1-46; 3) VH1-69; and 4) VH3-23. The resulting data sets were then analyzed using cross-donor phylogenetic analysis (Zhu et al., 2011).

The cross-donor phylogenetic analysis used here is similar to one we reported previously (Zhu et al., 2011). The major objective of cross-donor phylogenetic analysis is to bracket all phylogenetically similar sequences on a Neighbor-Joining (NJ) tree using known neutralizing antibody sequences. For the analysis here, we used pairs of neutralizing antibody sequences discovered from each of the six heavy chain 454 data sets (for a total of twelve neutralizing antibody sequences). The variable domain sequences of the following twelve neutralizing antibodies were used: 1) VRC01 (donor 45); 2) VRC03 (donor 45); 3) $gVRC_{dC38}H3$ (donor C38); 4)VRC18.02 (donorC38); 5) 1B2530 (donor RU01); 6) 1NC3 (donor RU01); 7) 8ANC131 (donor RU08); 8) 8ANC134 (donor RU08); 9) VRC13.01 (donor 44); 10) VRC13.02 (donor 44); 11) VRC16.01 (donor C38); and 12) VRC16.02 (donor C38).

The method was carried out in an iterative fashion and began by randomly shuffling all the sequences in a data set. Shuffling of the sequences removed any potential bias in the ordering of the sequences and improved convergence (Zhu et al., 2013). After sequence shuffling, the data set were split into FASTA files each containing up to 5000 sequences. All the neutralizing antibody sequences along with a germline gene sequence were added to each FASTA file. The germline gene sequence was chosen based on the VH family germline gene filter and was used as the outgroup in the NJ tree. A multiple sequence alignment for each FASTA file was generated using CLUSTAL0 (Larkin et al., 2007) with default settings. Each multiple sequence alignment was then converted into a distance matrix using the ***dnadist*** program (with default settings). From this, a NJ tree was constructed using the ***neighbor*** program (with default settings). Both ***dnadist*** and ***neighbor*** are part of the PHYLIP package v3.69 and can be downloaded from: (http://evolution.genetics.washington.edu/phylip.html).

Donor sequences were extracted from each NJ tree using a pair of neutralizing antibody sequences with matching VH genes as references. All donor sequences that were one node above the minimal-spanning tree containing the pair of neutralizing sequences were extracted from the NJ tree. The extracted sequences were then split into FASTA files containing no more than 5000 sequences and the above process repeated five times: 1) random shuffling of sequences; 2) addition of twelve neutralizing antibody sequences (including germline sequence); 3) alignment using CLUSTAL0; 4) conversion to a distance matrix using ***dnadist***; 5) construction of a NJ tree using ***neighbor***; and 6) extracting sequences one node above the minimal-spanning tree containing the pair of neutralizing antibody sequences. Sequences from each of the 454 data sets were pooled and sequence identities were computed between extracted donor sequences and the reference neutralizing antibody sequences. All sequences above 80% sequence identity were considered cross-donor relatives and the total number should appear off diagonal in Figure 4B.

Structure Modelling on the Mature Closed HIV-1 Viral Spike

The BG505.664 SOSIP crystal structure (PDB ID: 4TVP) was used to analyze the recognition of CD4-binding site by antibodies in a trimer context. First the trimer coordinates was transformed so that the trimer axis is on the z-axis and the center of the CD4-bindig site on outer domain of one protomer gp120 is on the x-axis. The outer domains of gp120 proteins of CD4-binding site antibody complexes were then superimposed to the x-axis-aligned protomer by aligning residues 252-392, 412-422, 437-476. The centers for the variable and constant domains of antibody Fab were defined by the averaged coordinates of Cα-atoms of the 2 pairs of conserved cystines in each domain, respectively. The long axis of antibody Fab was represent by the line connecting the centers (Figure S3). Approaching mode to the targeted gp120 protomer for each antibody was represented by the vector from the center of CD4-binding site on outer domain to the center of Fv domain of each antibody (Figure 2B). On the viral spike, the mode of recognition for CD4-binding site antibodies was

quantified with two angles, the latitudinal angle which explores the freedom between the viral membrane and the host cell membrane along the trimer axis, and the longitudinal angle which explores the freedom between gp120/gp41 protomers. The latitudinal angle was defined as the angle formed by the trimer axis (z-axis) and the projection of antibody vector on the y-z plane; and the longitudinal angle was defined as the angle formed by the protomer axis (x-axis) and the projection of antibody vector on the x-y plane.

Calculation of Positional Difference between gp120-Bound Antibodies

To obtain a common reference frame for comparison, all antibody-gp120 complexes to be analyzed were first superposed over the outer domain of gp120 (residue ranges: 252-392, 412-422, 437-476). The calculations of rotation angles and Cα-root mean square deviation (rmsd) between gp120-bound antibodies (or their heavy or light chains or structural components) were then carried out with the gp120-aligned structures. For comparison of heavy chains positions, the frame work regions (residues 1-22, 36-49, 76-82 and 103-114) were superimposed. For comparison of CDR L3s positions, the light chain residues 89-97 were superimposed. The superposition procedures were performed with the Superpose Molecules module in CCP4 (Collaborative Computational Project, 1994). The CHI angle of the CCP4 output was taken as the rotation angle between a compared pair.

Determination of the CD4 Supersite by Antibody Breadth

Average buried surface area (BSA) on gp120 was calculated for residues with BSA > 1Å^2 for the 16 gp120-antibody complexes, and the corresponding antibody neutralization potencies were averaged for each of those residues based on data from neutralization assays. Spearman correlation between BSA on gp120 and antibody potencies was calculated for BSA cutoffs = 0 to 85 Å^2 and potency logIC_{50} cutoffs = 0.60 to 1.62 (μg/ml).

Analysis of Antibody Paratope Chemistry

Representative structures of gp120-antibody complexes, including 11 from 2 VH gene-restricted classes (VH1-2 and VH1-46) and 4 CDR H3-dominated classes, were used in the analysis. Residues from each antibody were defined as part of the paratope if greater than 30% of residue surface area was buried in the interaction with gp120 in the relevant complex crystal structure. Buried surface area was calculated with SurfV (Nicholls et al., 1991). Paratope residues engaged in hydrogen-bonded or salt bridge interactions with gp120 were defined using hbplus (McDonald and Thornton, 1994).

Definition of Protein Atom Types

The atoms in twenty natural amino acids are divided into 30 protein atom types and are classified into 5 atom type groups (Table S7).

gp120 Proteoliposomes

A proteoliposome platform (Lingwood et al., 2012) was used to display multivalent gp120 for BCR activation. One gram of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) and 1,2-dioleoyl-sn-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid) succinyl] (nickel salt) [DGS-NTA(Ni)] (Avanti Polar Lipids Inc.) were mixed a 1:1 molar ratio was evaporated under a stream of nitrogen for 1 hr. The dry lipid film was then rehydrated in 1000 ml of liposome buffer [50 mM HEPES, 150 mM NaCl, pH 7.25 (HBS)] and shaken for 40 min, all the time being heated above the Tm of the lipid mixture. The resulting homogeneous suspension was subjected to ten freeze-thaw cycles and then extruded 21 times through a 100-nm pore polycarbonate membrane to produce uniform 100 nm liposomes (Avanti mini-extruder, Avanti Polar Lipids Inc.). An AviHis-tagged version of the YU2 gp120 core(Wu et al., 1996) was purified as wild type or mutant for recognition of its CD4 supersite (A281R, G366R, D368R, P369R) on a 17b affinity column as described(Kwon et al., 2012). One hundred twenty five micrograms of YU2 (wild type or mutant) was added to 0.5 mg of liposome suspension and left rocking at room temperature in a volume of 260 μl (HBS) for 45 min. The sample was then adjusted to 15% iodixanol (in 1.25 mL HBS) and overlaid with 1.75 mL, 0.5 mL and 0.5 mL of 10%, 2.5% and 0% iodixanol in HBS, respectively. Samples were then centrifuged for 200000g in a TH660 rotor (Sorvall) for 2 hr. The proteoliposome fraction, which concentrated at the 2.5%-0% iodixanol interface, was collected and dialyzed overnight (Slide-A-Lyzer Dialysis Cassette, 10000 MWCO, Cat# 66380, Thermo Scientific) to remove density gradient material. Following dialysis, proteoliposomes were pelleted (200000g, 2hr, TH660 rotor), resuspended in HBS and measured by the Pierce® BCA protein assay (Cat# 23227, Thermo Scientific). WT and mutant YU2 showed equal liposomal loading as measured by protein yield.

BCR Activation Assay

Mature and germline (gHgL) BCRs for VRC01 and VRC13 were stably expressed by lentiviral transfection (FEEKW vector (Luo et al., 2009)) of light chain and IgM heavy chain into a surface IgM negative clone of Ramos B lymphocytes (Lingwood et al., 2012). BCR positive cells [defined as staining positive for both light chain (mouse PE-anti-human lambda chain, Cat# 555797, BD Biosciences) and IgM (APC-anti-human IgM, Cat# 314510, BioLegend)] were then sorted by flow cytometry (BD FACSAria, BD Biosciences) and amplified. Cells were evaluated for triggering when BCR reached >85% expression. For signaling, 1×10^6 cells expressing mature or germline BCR were exposed to 2.5 μg/ml and 5 μg/ml YU2 proteoliposome, respectively. The kinetics of calcium flux in response to BCR stimulation was measured by flow cytometry as the ratio of the Ca^{2+} bound/unbound states of the dye Fura Red (Novak and Rabinovitch, 1994). Ratiometric measures for individual cells were averaged, smoothened (Kinetic analysis, Flow Jo software) and standardized to total Ca^{2+} flux as measured by exposure of cells to 10 μg/ml ionomycin. Comparable

surface expression between all four BCRs was confirmed by equivalent Ca^{2+} flux following exposure to mouse 5 μg/ml anti-human IgM $F(ab')_2$ (Cat# 9023-01, SouthernBiotech).

Table S2. Crystallographic Data Collection and Refinement Statistics, Related to Figure 1

Complex (antibody-gp120)	VRC13-93TH057	VRC16-93TH057	VRC18-93TH057	VRC27-Q32.17
PDB ID	4YDJ	4YDK	4YDL	4YDI
Data collection				
Space group	$P2_12_12$	C2	$P2_1$	C222
Cell dimensions				
a, *b*, *c* (Å)	293.8, 67.1, 93.6	125.8, 109.7, 100.1	66.3 213.1 68.2	141.7, 172.3, 91.8
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 126.3, 90.0	90.0 90.2 90.0	90.0, 90.0, 90.0
Resolution (Å)	50.00-2.31 (2.34-2.30)	50.00 -2.05 (2.09-2.05)	50.00-1.80 (1.83-1.80)	50.00-3.45 (3.57-3.45)
R_{sym} or R_{merge}	0.141(0.833)	0.072(0.462)	0.063(0.721)	0.143(0.348)
$I/\sigma I$	13.4(2.0)	22.6(2.2)	17.6(2.0)	10.8(2.1)
Completeness (%)	99.3(95.2)	98.5(88.5)	99.0(98.9)	88.6(58.1)
Redundancy	5.0(3.2)	3.7(3.1)	3.5(3.2)	5.5(3.4)
Refinement				
Resolution (Å)	2.30	2.05	1.80	3.45
No. reflections	81465	67455	175574	13713
R_{work}/ R_{free}	22.3/25.0	16.6/20.6	17.4/20.7	22.2/27.6
No. atoms				
Protein	11354	6071	12050	5959
Ligand/ion	367	280	353	215
Water	375	462	1402	25
B-factors				
Protein	90.3	52.2	43.9	108.8
Ligand/ion	88.5	83.4	66.7	112.50
Water	63.7	49.8	49.3	70.7
R.m.s deviations				
Bond lengths (Å)	0.003	0.003	0.007	0.004
Bond angles (°)	0.731	0.754	1.030	1.021
Ramachandran statistics				
Favored (%)	94.1	96.8	98.5	94.5
Outliers (%)	0.2	0.13	0	0.54

Complex (antibody-gp120)	8ANC131-YU2	8ANC134 –Q842.d12	1B2530-93TH057	HJ16-HT593.1
PDB ID	4RWY	4RX4	4YFL	4YE4
Data collection				
Space group	$P2_12_12_1$	H3	$P2_1$	I222
Cell dimensions				
a, *b*, *c* (Å)	66.2, 67.1, 214.1	220.4, 220.4, 118.5	67.0, 57.3, 254.7	87.7, 99.0, 180.4
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 120.0	90.0, 90.11, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50.00-2.39 (2.45-2.39, 2.16-2.12)	50.00-3.4 (3.58 -3.45)	50.00-3.43 (3.49-3.43)	50.00-2.71 (2.81-2.71)
R_{sym} or R_{merge}	0.093 (0.351, 0.477)	0.103 (0.775)	0.129 (0.488)	0.085(0.444)
$I/\sigma I$	12.36 (3.31, 1.74)	14.36 (2.34)	6.45 (2.21)	12.41 (2.32)
Completeness (%)	72.8 (53.8, 17.3)	99.6 (100)	92.1 (93.5)	94.5 (67.8)
Redundancy	6.2 (4.6, 2.8)	4.1 (4.1)	2.4 (2.1)	4.6 (2.5)
Refinement				
Resolution (Å)	28.84-2.13	26.90-3.45	28.78 -3.43	41.9-2.71
No. reflections	43292	28017	24702	20510
R_{work}/ R_{free}	21.5/24.6	20.7/24.6	27.0/32.4	21.7/26.1
No. atoms				
Protein	6149	11740	12254	6043
Ligand/ion	117	215	120	126
Water	228	0	0	37
B-factors				
Protein	58.8	106.20	209.0	69.7
Ligand/ion	73.3	121.80	212.2	103
Water	47.8	NA	0	47
R.m.s deviations				
Bond lengths (Å)	0.003	0.009	0.010	0.004
Bond angles (°)	0.783	1.36	1.853	0.82
Ramachandran statistics				
Favored (%)	96.3	95.9	90	93.2
Outliers (%)	0.39	0.88	1.1	1.1

Values in parenthesis denote highest resolution shell.

Table S3A. Basic Information of Monoclonal CD4-Binding Site Antibodies Isolated from Human Donor Samples, Related to Figure 2

Type (structural mode of HIV-1 recognition used by each antibody class)	**Class** (Clonal families with related ontogenies in separate donors)	**Clonal lineage** (Clonal related antibodies which derive from the same naïve rearranged B cell)	**Donor** (Subject from whom the clonal family was isolated)	**Representative monoclonal** (Best characterized monoclonal antibody of each clonal family)	**Isolation method**[a]	**VH gene**	**CDR H3 length**[a, b] (amino acids)	**Vλ/κ gene**[a]	**CDR L3 length**[a, b] (amino acids)	**Breadth on 172-virus panel**[a] (Defined as IC_{50}<50 µg/ml)
CDR H3 dominated	HJ16 class	HJ16	242315	**HJ16***	EBV	3-30	18	κ 4-1	8	30%
	VRC13 class	VRC13	44	**VRC13***	RSC3-probe	1-69	21	λ 2-14	6	82%
	VRC16 class	VRC16	C38	**VRC16***	RSC3-probe	3-23	20	κ 1-39	9	57%
	CH103 class	CH103	CHAVI 505	**CH103***, CH104, CH106	RSC3-probe	4-59	13	λ 3-1	10	56%
VH gene restricted	VRC01 class [VH1-2-derived]	VRC01	45	**VRC01***, VRC02, VRC03, VRC06, VRC07, VRC08, NIH45-46	RSC3-probe	1-2	12	κ 3-20	5	90%
		VRC23	127/C	**VRC23***	RSC3-probe	1-2	12	κ 3-15	5	65%
		VRC18	C38	**VRC18***	RSC3-probe	1-2	10	κ 3-20	5	81%
		VRC-CH31	CHAVI0219	VRC-CH30, **VRC-CH31***, VRC-CH32, VRC-CH33, VRC-CH34	RSC3-probe	1-2	13	κ 1-33	5	81%
		VRC-PG20	IAVI23	VRC-PG19, VRC-PG19b, **VRC-PG20***, VRC-PG20b	RSC3-probe	1-2	13	λ 2-14	5	78%
		12A21	IAVI57	**12A21***, 12A12, 12A30	2CC-probe	1-2	13	κ 1-33	5	85%
		VRC-PG04	IAVI74	**VRC-PG04***, VRC-PG04b	RSC3-probe	1-2	14	κ 3-20	5	79%
		VRC27	Z258	**VRC27***	RSC3-probe	1-2	14	κ 1-33	5	78%
		3BNC117	RU3	**3BNC117***, 3BNC60, 3BNC62	2CC-probe	1-2	10	κ 1-33	5	84%
	8ANC131 class [VH1-46-derived]	1B2530	RU1	**1B2530***, 1NC3, 1NC7	2CC-probe	1-46	16	λ 1-47	11	43%
		8ANC131	RU8	**8ANC131***, **8ANC134***, 8ANC37	2CC-probe	1-46	16	κ 3-20	9	79%

*Structural analysis of this monoclonal antibody is described in this paper.
[a]Pertains to monoclonal antibodies marked with asterisks.
[b]Kabat definition, in which CDR H3 is 2 amino acids shorter than the IMGT definition.

Table S3B. Characteristic Features of Antibody Types Targeting the CD4 Supersite, Related to Figure 2

	CDR H3-dominated	**VH-gene-restricted**	
Observed classes	CH103, HJ16, VRC13, VRC16	VRC01-class (VH1-2)	8ANC131-class (VH1-46)
Overall	All antibodies have the same latitudinal binding orientation as CD4, but freedom in longitudinal dimension	All antibodies have essentially the same binding orientation	All antibodies have essentially the same binding orientation, but more freedom than VH1-2
Heavy chain	CDR H3 > 50% contact surface	CDR H2 > 50% contact surface	CDR H2 > 50% contact surface
Light chain	Diverse CDR L2 and L3 interaction modes.	5-amino-acid CDR L3 with third position hydrophobic and fourth position E/Q	CDR L3 length not constrained to 5 amino acids.

Table S4. Contribution of Antibody CDRs to HIV-1 gp120 Binding, Related to Figure 2

Antibody	Light chain paratope (Å^2)				Heavy chain paratope (Å^2)			
	CDRL1	CDRL2	CDRL3	Total	CDRH1	CDRH2	CDRH3	Total
CH103	72.1	116.6	74.42	362.1	30.1	77.3	**328.9**	461.8
HJ16	146.7	0	215.9	362.6	31.9	120.5	**482.1**	634.5
VRC13	7.01	14.8	0	21.8	114.9	23.1	**574.0**	966.1
VRC16	106.8	40.8	110.01	297.0	0	191.0	**674.1**	944.9
VRC01	113.4	0	113.4	308.6	2.2	**607.2**	122.5	883.4
VRC04	58.7	0	131.2	216.8	19.6	**431.6**	181.0	749.9
VRC18	54.7	0	121.1	220.0	65.0	**637.8**	94.7	920.0
VRC20	57.1	0	149.2	206.3	34.7	**751.6**	137.6	1111.2
VRC23	24.6	0	182.0	342.8	10.5	**567.6**	135.1	876.3
VRC27	13.3	0	133.2	215.4	24.8	**569.3**	158.1	991.6
12A21	120.0	0	84.4	245.3	4.4	**698.6**	229.7	1131.3
3NBC117	56.6	0	97.6	269.2	53.4	**484.8**	107.6	802.4
CH31	118.4	0	168.6	395.8	39.8	**425.6**	100.5	618.0
1B2530	1.6	0	216.8	245.2	0	**553.1**	121.5	820.8
8ANC131	19.8	11.2	158.2	212.9	24.7	**464.2**	155.2	992.8
8ANC134	13.6	0	0	180.0	25.6	**448.6**	181.8	1044.0

Table S5. 454 Pyrosequencing of B Cell Transcripts by Donor and by VH-Gene Family, Related to Figure 4

Donor	45[a]	C38[b]	RU01[c]	RU08[c]	44[c]	C38[c]
Number Raw Reads	696,971	290,744	345,080	315,711	478,517	395,545
Number Reads > 300bp	696,971	266,975	325,054	296,274	457,961	368,902
Number Reads (VH1-2)	175,137	77,797	17,468	16,108	41,384	0
Number Reads (VH1-46)	41,535	32,959	18,972	4,053	43,793	0
Number Reads (VH1-69)	220,497	56,986	198,536	187,177	213,058	265
Number Reads (VH3-23)	0	0	0	0	0	59,236

[a]. Donor 45 data was from a previous deposit at the National Center for Biotechnology Information Short Reads Archives under accession code SRP052625 (Wu *et al.*, 2015). The heavy chain data from year 2008 time point was used in this study.
[b]. Donor 38 data (for VRC18.02) was from a previous deposit at the National Center for Biotechnology Information Short Reads Archives under accession code SRP026397 (Zhu *et al.*, 2013).
[c]. NGS data for donors RU01, RU08, 44 and C38 (for VRC16.01) were obtained in this study and deposit at the National Center for Biotechnology Information Short Reads Archives under accession code SRP055520.

Table S7. The Atom Types of 20 Natural Amino Acids, Related to Figure 7

Atom type Group *	Polar atom Group	Protein Atom Type	Atom Radius (Å)	Description	Number
B	N	NH1	1.65	Backbone NH	1
B		CB	1.76	Backbone C	2
B		CH1E	1.87	Backbone CA (exclude Gly)	3
B	O	OB	1.40	Backbone O	4
C		CH2G	1.87	Gly CA	13
C		CH0	1.76	Arg CZ, Asn CG, Asp CG, Gln CD, Glu CD	5
C		CH1S	1.87	Sidechain CH1: Ile CB Leu CG, Thr CB, Val CB	6
C		CH2P	1.87	Pro CB, CG and CD	14
C		CH2E	1.87	Tetrahedral CH2 (except CH2P and CH2G) All CB	7
C		CH3E	1.87	Tetrahedral CH3	8
A		CR1E	1.76	Aromatic CH (except CR1W, CRHH, CR1H)	9
A		CF	1.76	Phe CG	21
A		CY2	1.76	Tyr CZ	19
A		CY	1.76	Tyr CG	23
A		CR1W	1.76	Trp CZ2 and CH2	18
A		CW	1.76	Trp CD2 and CE2	24
A		C5W	1.76	Trp CG	30
C		CRHH	1.76	His CE1	25
C		CR1H	1.76	His CD2	27
C		C5	1.76	His CG	28
S		SC	1.85	Cys S	20
S		SM	1.85	Met S	22
P	OH1	OH1	1.40	Alcohol OH (Ser OG, Thr OG1 and Tyr OH)	10
P	OCS	OS	1.40	Sidechain O. Asn OD1 and Gln OE1	12
P	OCS	OC	1.40	Carboxyl O (Asp OD1, OD2 and Glu OE1, OE2)	11
P		ProN	1.65	Pro N	29
P	NS	NH1S	1.65	Sidechain NH: Arg NE, His ND1, NE2 and Trp NE1	15
P	NS	NH2	1.65	Asn ND2 and Gln NE2	17
P	NS	NC2	1.65	Arg NH1 and NH2	16
P	NS	NH3	1.50	Lys NZ	26

*: First column from left: B is for backbone atom types; C is for aliphatic carbon atom types; A is for aromatic carbon atom types; S is for sulfur atom types; P is for polar atom types. Second column from left: The polar atom groups. N denotes backbone N; O denotes backbone O; OH1 denotes OH1 atom type; OCS denotes OS and OC atom types; NS denotes NH1S, NH2, NC2 and NH3 atom types. The third and fourth columns show the protein atom types and van der Waals radii respectively. The description of the protein atom types are shown in the fifth column.

SUPPLEMENTAL REFERENCES

Arnaout, R., Lee, W., Cahill, P., Honan, T., Sparrow, T., Weiand, M., Nusbaum, C., Rajewsky, K., and Koralov, S.B. (2011). High-resolution description of antibody heavy-chain repertoires in humans. PLoS One *6*, e22365.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res *32*, W615-619.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).

Doria-Rose, N.A., Schramm, C.A., Gorman, J., Moore, P.L., Bhiman, J.N., DeKosky, B.J., Ernandes, M.J., Georgiev, I.S., Kim, H.J., Pancera, M.*, et al.* (2014). Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature *509*, 55-62.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. Journal of molecular biology *372*, 774-797.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R.*, et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Lingwood, D., McTamney, P.M., Yassine, H.M., Whittle, J.R., Guo, X., Boyington, J.C., Wei, C.J., and Nabel, G.J. (2012). Structural and genetic basis for development of broadly neutralizing influenza antibodies. Nature *489*, 566-570.

Luo, X.M., Maarschalk, E., O'Connell, R.M., Wang, P., Yang, L., and Baltimore, D. (2009). Engineering human hematopoietic stem/progenitor cells to produce a broadly neutralizing anti-HIV antibody after in vitro maturation to human B lymphocytes. Blood *113*, 1422-1431.

McDonald, I.K., and Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. Journal of molecular biology *238*, 777-793.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281-296.

Novak, E.J., and Rabinovitch, P.S. (1994). Improved sensitivity in flow cytometric intracellular ionized calcium measurement using fluo-3/Fura Red fluorescence ratios. Cytometry *17*, 135-141.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol *276*, 307-326.

Tramontano, A., Chothia, C., and Lesk, A.M. (1990). Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. J Mol Biol *215*, 175-182.

Wu, L., Gerard, N.P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A.A., Desjardin, E., Newman, W*., et al.* (1996). CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. Nature *384*, 179-183.

Zhu, Z., Qin, H.R., Chen, W., Zhao, Q., Shen, X., Schutte, R., Wang, Y., Ofek, G., Streaker, E., Prabakaran, P*., et al.* (2011). Cross-reactive HIV-1-neutralizing human monoclonal antibodies identified from a patient with 2F5-like antibodies. Journal of virology *85*, 11401-11408.